



## King's Research Portal

DOI:

[10.1080/09500782.2018.1429463](https://doi.org/10.1080/09500782.2018.1429463)

*Document Version*

Peer reviewed version

[Link to publication record in King's Research Portal](#)

*Citation for published version (APA):*

Schissel, J. L., Leung, C., López-Gopar, M., & Davis, J. R. (2018). Multilingual learners in language assessment: assessment design for linguistically diverse communities. *Language and Education*, 167-182.

<https://doi.org/10.1080/09500782.2018.1429463>

### **Citing this paper**

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

### **General rights**

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

### **Take down policy**

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

## **Multilingual learners in language assessment: Assessment design for linguistically diverse communities**

### **Abstract**

The assessments designed for and analyzed in this study used a task-based language design template rooted in theories of language reflecting heteroglossic language practices and funds of knowledge learning theories, which were understood as transforming classroom teaching, learning, and assessment through continua of biliteracy lenses. Using a participatory action research model, we created assessment instruments for pre-service English teachers in Oaxaca, Mexico, integrating language practices from communities and classrooms into assessments. Participants completed two reading and writing tasks. Task one was intentionally designed to engage learners' English and Spanish languages resources. Task two was restricted to English-only. Our analyses indicated (1) that pre-service English teachers performed better on the multilingual task than the monolingual English task at a level of statistical significance and (2) that integrating multilingual resources within assessment design can allow test-takers to demonstrate more complex or high-order thinking skills in the language they are learning. We are offering some empirical evidence of an assessment approach that is consistent with the broadly supported principle of making use of all students' linguistic resources for the purpose of teaching and learning.

**Keywords:** Funds of knowledge; Heteroglossia; Multilingual learners; Participatory action research; Task-based language assessment

This project presents approaches to language assessment that valorize learners' multilingual language resources and addresses the long-term possibility of developing language assessments through collaborations with teachers and learners in linguistically diverse communities. Our project has been informed by, *inter alia*, orientations within Educational Linguistics (Hornberger 2001)<sup>1</sup> where our identification of the research problem and design of the study drew from relevant disciplinary histories, methods, and knowledge from education, linguistics, and other areas. Research in Educational Linguistics is problem-oriented and transdisciplinary, merging multiple perspectives to address a particular issue with novel approaches (Hult 2007; Spolsky 1978), drawing from theory, research, policy, and practice—with a particular emphasis on practice (Hornberger 2001). In situating our study within this field,

---

<sup>1</sup> This paper was presented as part of a symposium title “Educational Linguistics at 40: Past, Present, and Future” at the annual meeting of the American Association of Applied Linguistics, 2017.

we strive to engage in research, such as this project, where we can actively account for the social implications of our work.

This study is a part of a larger multilingualism project exploring an approach that would encourage learners to make full(er) use of their language resources within classroom-based language assessments. To accomplish this, we integrated heteroglossic language perspectives, which view dynamic language practices as routine and unexceptional (Bakhtin 1981), with funds of knowledge learning theories (González, Moll, and Amanti 2006). Using continua of biliteracy lenses (Hornberger and Link 2012), we have woven together theories and pedagogies of translanguaging and translingual practices with task-based language assessment approaches. These approaches have their foundations in Dewey (1938/1997), later taken up by Mohan (1986, 1990). Dewey emphasized social interaction within task-based design and in performing a language learning task. In emphasizing the social dimension of language tasks, we shift the traditional perspective that regards languages as entities with fixed boundaries to a view that sees purposeful language use as manifestation of the language practices of a community that may be multilingual. This approach has powerful implications for language assessment, and is responsive to calls by Shohamy (2011) and Otheguy, García, and Reid (2015) to embrace the language resources of learners in assessments.

Heteroglossic and funds of knowledge perspectives would suggest that the pre-service English teachers, who are learners and users of English participating in our study, will perform better on the multilingual task because such a task gives learners an opportunity to demonstrate/access a wider range of available language resources. At the university where we conducted our research in Oaxaca, Mexico, pre-service English teachers varied in their English language education backgrounds. The program prepares most of the English teachers for K-16

classrooms for the state of Oaxaca, but does not have an English proficiency level prerequisite for admissions. The pre-service English teachers at this university reflect the diverse range of experiences and English language education in Oaxaca rather than selecting for those with experiences limited to elite language academies.

In this paper, we investigate: *How do learners' performances on assessments that use multilingual or monolingual approaches compare?* We developed and administered assessments to pre-service English teachers in Oaxaca, Mexico ( $n = 39$ ). Findings of task performance involving multilingual materials can help us generate some of the empirical support to explore the ways in which this approach assesses learners' language skills. Our initial analyses indicated (1) that the pre-service English teachers performed better on the task that was accompanied by multilingual reading material than the task that only allowed monolingual English reading material at a level of statistical significance and (2) that integrating multilingual resources within assessment design can allow test-takers to demonstrate more complex or high-order thinking in writing in the language they are learning.

### **Conceptual framework: Developing assessment approaches for linguistically diverse communities**

In this section we will provide a conceptual discussion that invokes a heteroglossic view of language and funds of knowledge oriented learning theories. Continua of biliteracy lenses spotlight the additive facilitation of translanguageing and translingual pedagogies for learners, and task-based language assessment serves as a methodological device where these theories and pedagogies converge as key considerations for assessment design.

### **Theories of language and learning**

Blackledge and Creese (2010) have theorized how heteroglossia, first explored in the

work of Bakhtin (1981), takes dynamic, fluid communication among multilingual interlocutors as unremarkably normal, as part of the “natural” language ecology of a community. This perspective prioritizes a language perspective that moves beyond socially constructed definitions of *languages* as discrete units (Makoni and Pennycook 2005) to a holistic perspective where fluid and dynamic uses of all available linguistic resources are valorized. Heteroglossic language practices refer to the ways in which language/s, dialects and other semiotic resources are combined and used fluidly in communication in specific local contexts. These languages practices have been termed variously as *linguaging* (Lado 1979; Swain 2006), *translanguaging* (Blackledge and Creese 2010; García 2009; Hornberger and Link 2012), *translingual practices* (Canagarajah 2013) and *polylingual linguaging* (Jørgensen 2008; Jørgensen, Karrebaek, Madsen, and Møller 2011), among others. Despite the sometimes very subtle, conceptual and theoretical differences that underlie these terms, a common thread that ties them together is their rejection of the privileging the notion of languages as bounded and separate entities and native speaker idealisms, and instead valorize the fluid and dynamic use of language resources in local contexts.

Funds of knowledge, as an approach to education, positions community spaces as important areas where knowledge is developed (González, Moll, and Amanti 2006). Learning theories inspired by funds of knowledge have also taken account of linguistic diversity since the early 1990s (Moll, Amani, Neff, and González 1992) and are used in this framework to integrate local language practices within classroom assessments. In moving beyond views that merely recognize students’ backgrounds, literacies, practices from the home within schools, funds of knowledge approaches seek to transform learning experiences in schools by sustaining and incorporating (e.g., Paris 2012) lived experiences of learners. By honoring and incorporating learners’ knowledge that is traditionally excluded in the classroom, we work with an assessment

framework that integrates “reflection and action upon the world in order to transform it” (Freire 2000, 51). With this perspective on heteroglossia and funds of knowledge, we examine the possibilities of developing language assessments that reflect and exploit community-based multilingualism and community-engendered communication skills.

## **Pedagogies**

Translanguaging and translingual practices have garnered support as pedagogical practices. Translanguaging is associated with approaches that leverage learners’ language resources in the classroom and has largely looked at how to integrate these approaches in K-12 bilingual education (e.g., English as a second language, dual language) classrooms in the United States (de la Luz Reyes 2012; Palmer, Mateus, Martínez, and Henderson 2014; García and Kleyn 2016) and internationally (Blackledge and Creese 2010; Dodson 1985; Lewis, Jones, and Baker 2012; Lin and Wu 2015; Vaish and Subhan 2015). Translingual practices often highlight renegotiated power relationships within writing, in particular how an author views their position or power in relation to the self and others (Canagarajah 2013) have gained attention in composition and rhetoric studies in the United States and often emphasize on multilingual writers in higher education contexts (Horner, Lu, Royster, and Trimbur 2011; Jain 2014), though translingual practices have been applied in other contexts as well (Higgins, Nettell, Furukawa, and Sakoda 2012; Pacheco and Smith 2015).

In looking at classroom practices, we position the continua of biliteracy as a tool for organizing pedagogical approaches that sustain the lived experiences of learners in the classroom. Building on decades of research on the continua of biliteracy (e.g., Hornberger 1989, 2003; Hornberger and Skilton-Sylvester 2000), Hornberger and Link (2012) explained the utility of the continua of biliteracy to

offer a set of lenses through which educators can organize pedagogy in today's linguistically diverse classrooms, because the continua bring into focus all the dimensions—of context, content, media, and development—that research says should be taken into account in creating a learning environment that recognizes and builds on the language and literacy repertoires students bring to school. (243)

The continua of biliteracy lens anticipates and extends translanguaging by drawing attention to the multidimensional aspects that contribute to the development of biliteracy (Baker 2003) in preparation for different linguistic situations inside and outside of educational settings. The different continua of context, content, media, and development and how they are further delineated within the continua of biliteracy integrate the “fluid, multilingual, oral, contextualized practices at the local level, [which] are essential for learners’ development” (276), and which we also see as essential for the assessment of learners’ development. We view the contributions by these scholars as foundational in motivating and moving forward innovations from communities and classrooms *to* assessments.

### **Assessments**

We view assessment practices as an integral part of a complex of educational theories and socio-political values (Shohamy 2006). Shohamy (2011) has called for the development of assessments that recognizes the language practices of learners as an issue of social justice and educational equity. In multilingual societies “...by conducting language tests in a given language, messages are being transmitted regarding the priority of dominant languages while marginalizing others” (421). In a similar vein, Otheguy, García, and Reid (2015) have called into question the validity of assessment scores that isolate languages. They argue that “if schools want to test students’ linguistic ability – their ability to do these things with language – it doesn’t

make sense to ask them to perform using only some of their linguistic repertoire” (300). These concerns urge researchers to consider the possibilities of assessment methods to engage with learners’ total language resources in assessments.

Some assessment adaptations designed to provide a more level playing field such as test accommodations, or changes to the test response have attempted to integrate multilingual skills. These methods include using a word-for-word translation dictionary or bilingual test forms. Yet these attempts at accommodation promote a standpoint that views learners’ other languages as at best irrelevant and worst as a potential threat to measurement or interference with the test construct (Author 2014, 2015; Rea-Dickens et al., 2011; Stansfield 2011). The design principles of task-based language assessments (TBLA) extended from Mohan (1986, 1990) are uniquely positioned to incorporate multilingualism while also viewing multilingualism as an asset to the assessment process or instrument.

Task-based language assessments derive from task-based language teaching (TBLT). Pica (2005) writes that TBLT is pedagogically oriented and driven by theories of language or language learning. Ellis (2003) further elaborates that TBLT focuses on meaning making, allows students to use their own resources, and has a clearly defined goal of task completion by filling an information gap rather than using the language “correctly.” In extending TBLT into TBLA, Van Gorp and Deygers (2014) emphasize that the assessments measure “what learners are able to *do* with language as opposed to what they *know* about language” (1, emphasis in original). This approach to language assessment connects with the direct testing approach, where development of the assessment mirrors meaningful, authentic, real-world language use (Bachman and Palmer 2010; Van Gorp and Deygers 2014). In describing the strengths of TBLA, Norris (2016) included that this approach to assessment enabled predictions about the extent to which learners



could use the language rather than merely know something about its rules and vocabulary. We are extending these ideas by incorporating learners' multilingualism as an important part of real-world language use.

We argue that the tenets of an expanded approach of TBLA that focus on meaning making and authenticity language use in context are consistent with our approach that values the learners' language resources. We are offering some empirical evidence that supports the broadly supported principle of making use of all linguistic resources at students' disposal which are potentially applicable to many linguistically diverse communities. Furthermore, Shohamy (2011) and Otheguy, García, and Reid (2015) view monolingual biases in language assessments as key equity issues facing linguistically and culturally diverse students across the globe because “[m]any language assessment and teaching practices work to the detriment of bilingual children worldwide” (Otheguy, García, and Reid 2015, 283) because they are not allowed to perform to the best of their potentials. The project combines theories and research to create multilingual assessment approaches that can work to *benefit* linguistically diverse children and adults.

### **Study design and data collection**

Our study used participatory action research (PAR) methods in shaping the choices for the design, administration of the assessment instruments, as well as the methods for data collection and analysis. PAR is research *with* participants, rather than *for* (Whyte, 1991). Within these methods is the ontological assumption that involvement of participants in all key aspects of the research process is crucial because their knowledge is intrinsic to the production of the work (Borda, 2001; McIntyre, 2008). Engaging in participatory methodology is essential to construct multilingual assessment approaches with scores inferences guided by teacher and learner purposes. In translating PAR into practice in our research project, the research team worked

closely with teacher-collaborator Mr. Julio Morales<sup>2</sup> to understand the purposes of the classroom language assessments and incorporate the lived experiences of learners within the assessment.

PAR methodology also drew from the extensive experiences of the research team and worked to make the participants more visible. Author 3, for example, is a member of the local and university community. Since 2015 Author 1 has been and continues to engage in work with the community. The research team employed Ibrahim's (2008) critical ethnographic approach of "hanging out" methodology. For our project hanging out—or engaging in causal settings with instructors and current and former students from the university such as classes, dinners, presentations, and other social gatherings—provided rich information about the current needs and desires of the community. As a member of the local community, Author 3 regularly engages in such activities. During the design and data collection for this project, Author 1 spent ten days immersed in such activities, while also maintaining regular contact with Mr. Morales before and after the ten-day stay. In negotiating multilingual approaches to language assessment for this study, the research team has continued to work closely with Mr. Morales; and the findings from this project have informed our continued collaboration.

### **Community context**

In Mexico, language teachers and learners must manage increased scrutiny from the federal government. This means, *inter alia*, pre-service English teachers having to take the large-scale, standardized English proficiency test TOEFL (Test of English as a Foreign Language) for both accreditation purposes for universities and graduate requirements for students (British Council 2015; Vazquez, Guzman, and Roux 2013). The increased attention at the federal level in the country focused on ascertaining individual teacher language proficiency through the use of

---

<sup>2</sup> Not a pseudonym.

large-scale standardized assessments assumes that assessing English teachers' proficiency would lead to effective teaching practice.

This assumption about the relationship between TOEFL scores and professional performance distracts attention away from teachers and students' use of their language resources in the classroom and the implications for classroom-based language assessments. For our project, we have been working with a university in Oaxaca, Mexico. At this university, initiatives focused on large-scale, standardized assessments in Oaxaca (Ramírez Victoria and León Jiménez 2012) have also fostered interest in understanding the role of assessment in education generally, including expanding the potential of classroom-based language assessments and our research project. For the English language program at this university, pre-service teachers are administered assessments three times a semester during an eight-semester undergraduate program. The courses each semester comprise three thematic units, with learning objectives created to address the domains of reading, writing, speaking, listening, vocabulary, and grammar. Assessments are based on these learning objectives for those different domains.

As a region, Oaxaca is uniquely situated to provide insights in the theory and practice of multilingual teaching, learning, and assessment. Oaxaca is among the most linguistically and culturally diverse areas in the country, with numerous language families and 16 ethnic groups such as Zapotec, Mixtec, Chatino, Triqui, and Mixe communities among others. Since 2003 the Mexican government formally recognizes 364 Indigenous languages, which, among other uses, allows for their use in educational contexts. Author 3 (2013) has and continues to document the heteroglossic practices that are being promoted in local schools by finding ways to include the newly-legitimated Indigenous languages, as speakers of these languages remain largely stigmatized. Building on language teachers' foundational knowledge and experiences with

classroom language assessments, we have been working to expand their professional repertoire specifically with the multilingual local language practices of Oaxaca in mind.

### **Instrument design and administration**

The researcher team created assessment instruments in alignment with the course syllabus for a fifth semester English language course for pre-service English teachers in Oaxaca. Using the multilingual task-based assessment specifications (see Appendix A), the researchers designed two tasks to elicit responses offering ‘advice’ or ‘recommendations’. We aligned the assessment objectives and rubric criteria for the two tasks. Thus, it was also our intention to administer two tasks with aligned objectives and rubric criteria in order to compare task performance. The task topics were chosen based on current issues commonly covered in Mexican media (e.g., Cruz Martínez, 2016) and the course objectives as listed in the syllabus. Task one asked participants to provide advice to prevent the onset of type II diabetes, whereas task two asked participants to make a recommendation for a litter clean-up program for campus. The tasks asked participants to synthesize three readings to write a letter or email in English. There were three reading materials per task adapted from authentic sources. Task one had one reading in Spanish, and two in English. Task two had readings in English-only. Thus, task one required participants to use their bilingual skills for reading comprehension, while task two restricted participants to use only English.

Our use of asking the learners to produce responses in English was decided through discussions with the teacher-collaborator, Mr. Morales. In working with Mr. Morales, we discussed the purposes of the assessments and the types of information that he would find useful as evidence of student learning. He expressed a need to have a test that would solicit writing from learners for an audience that required the text in English. Yet he was amenable to including

languages other than English in the language inputs of the test. For this first use of this assessment approach, we focused on the common language, Spanish in addition to English, as not all students shared another language, i.e. Indigenous languages. For the language of the response, Mr. Morales requested for it to be completed in English. In working to integrate learners' language resources within the test, we also drew from Mr. Morales' description of classroom language practices and the research team's knowledge that it was a common practice in the classroom at this university in Oaxaca for learners use or seek out materials in Spanish to supplement their understandings. We then used the same assessment design procedures for the second task, but used readings only in English, which had been the common practice with previous assessments. Thus task one represented a closer match to the language practices that drew on learners' multilingual resources and that were common in classroom instruction, whereas the monolingual (English only) task two was familiar to learners previous assessment experiences with Mr. Morales<sup>3</sup>.

In December 2016 near the end of the semester, 44 pre-service English teachers in their 5<sup>th</sup> (n= 25) and 7<sup>th</sup> (n = 19) semesters of English language studies took the assessment. They were given 20 minutes per task, and were asked to complete an attitudinal survey after completing the assessment. All participants were given task one before task two. We wanted to begin with the task that mirrored classroom language practices, followed by the assessment that more closely aligned with previous assessments solely in English. 40 pre-service teachers completed both tasks. After rating, one participant's completed tasks were removed due to significant rater disagreement; thus for our analysis, n = 39. Assessment administrations were video recorded and Mr. Morales and Author 1 remained in the room while participants took the

---

<sup>3</sup> Member check October 11, 2017

assessments.

### **Rater training**

To adjudge the assessment performance, the tasks were double-rated using a binary *yes-no* checklist with six criteria, extended from the lenses of the continua of biliteracy. Assessment scores were calculated as a sum of the criteria where yes is given one point and no was given zero points for each test-taker (hence a scale of 0-6). The assessment objectives and six rubric criteria were aligned across the two tasks. Table 1 compares the six criteria for each task.

[insert Table 1 near here]

For each criterion, the rubrics were near identical, in particular for criteria one, two, and six. Differences for author voice were based on information from Mr. Morales. He stated that the pre-service English teachers in his classes varied in excelling in both or either types of formality, and thus he preferred having both formal and informal registers represented rather than only one form. For criterion four, which will be discussed further in our paper, we note that the skills of an analysis of cause and effect and evaluation are not identical, but both were designed to require a response that would discuss an effect and were scored for this evidence.

Eight raters were recruited from a graduate Teaching English to Speakers of Other Languages (TESOL) program in the United States. They had experience teaching in linguistically diverse communities between 1 and 20 years (mean: 4.25 years) and either were in the process of attaining or had attained higher degrees in TESOL. All raters were recruited from a pool of individuals who we consider to be bi-/multilingual, and given their professional experience and training with English language education were deemed sufficient for the rating of these assessment responses. They were offered a small compensation. Two raters declined the compensation.

The raters were divided into two groups to rate either task one or two. Each criterion required interpretation on the part of the rater. The raters needed to pass an initial rater training using three sample tasks, which lasted from 10-30 minutes and involved discussions about the rubric criteria. Raters engaged in discussions around guidelines for interpreting the rubric in relation to what types of response information was necessary and sufficient to score a *yes*. Rater G, for example, asked for clarification about the binary checklist rubric, noting that a response “doesn’t have to be a 100% to be a *yes*.” For context about Mexico, if the writer used the term “here” rather than “Mexico,” a *yes* score was given because it would be fair for the writer to assume that “here” would be understood as “Mexico.” Raters were instructed to provide explanations on the open-response portion of the rubric when additional clarification was necessary. After completing the training on either task one or task two, each rater scored a total of 20 tasks. Each rater scored 10 of the same tasks as two other raters (see Table 2). The numbers listed indicate the number assigned to the pre-service English teachers.

[insert Table 2 near here]

A member of the research team led the task rating in-person. The scores were entered using *Qualtrics*, an online survey program. Six scores per rater were compared to pre-rated scores calculated by the research team in order to monitor how raters were interpreting the rubric during the rating process. Rater scores were consistent with pre-rated scores, providing evidence of rater consistency in interpreting and applying the rubric.

### **Analyses and findings**

Findings represent support of integrating multilingual skills within a TBLA framework. As one of the first studies investigating a purposeful, integrated multilingual approach to language assessment, the information in the following sections document procedures necessary

to support such analyses and results. Rater reliability was calculated with R (2016) using *Krippendorff's Alpha*. We then ran a paired-sample T-test to determine if pre-service English teachers did significantly better on the multilingual task.

**Rater reliability.** Krippendorff's Alpha was chosen because it can be used to compute interrater reliability for ordinal data from any number of raters with either large or small sample sizes (Gwet 2016; Krippendorff 2011; Zapf, Castell, Morawietz, and Karch 2016). The R code for the Krippendorff calculation was provided by Zapf et al (2016) in their paper. According to Zapf, Castell, Morawietz, and Karch (2016), alpha of 0.70 to 0.88 indicates *substantial agreement between raters*. Krippendorff's Alpha is 0.764 for task one and 0.761 for task two with 95% CIs [0.59, 0.86] and [0.60, 0.86], respectively. These results support the use of raters' scores for analysis and comparison of task performance. To facilitate comparison between tasks, rating pairs were then averaged to provide a single score for each student on each task.

**Comparing task performance.** The data indicated that pre-service English teacher participant in Oaxaca ( $n = 39$ ) performed significantly better on an English writing assessment when they used readings that were in Spanish and English than on the assessment task that used readings only in English,  $t(38) = 3.32, p = 0.002$  (see Table 3). These findings provide initial empirical evidence to support continued development and use of this multilingual task-based assessment approach.

[insert Table 3 near here]

In comparing task performance, one additional finding was revealed upon closer examination of performance on the different rubric criteria. As described earlier, each criterion on the rubric analyzed a specific skill or strategy required for task completion. We initially reviewed task performance descriptively to look at how participants performed on each criterion



by looking at the average scores by criterion for task one and task two. We observed that criterion four exhibited the largest difference in score average (task one = 0.613; task two = 0.04). Because criterion four asked for more higher-order thinking compared with other criteria, we wanted to test the impact of criterion four to our initial result. When we ran the T-test without criterion four being included on either task, the differences between task performance were not statistically significant, suggesting that task four was driving our initial result. Table 4 presents a comparison of task performance with and without criterion four.

[insert Table 4 near here]

These findings suggest to us that we need to investigate further how the multilingual task design may facilitate a more successful engagement in complex tasks by learners, which holds implications for language teaching and learning. We elaborate on our findings in the discussion section.

### **Caveats and limitations**

The study was limited in terms of context-specific constraints. The small sample size, though large for the context, impedes generalizability to other contexts. There were also time constraints from several students arriving late to class. Our participants were drawn from two different courses: 5<sup>th</sup> and 7<sup>th</sup> semester. This limitation was seen as less significant because Mr. Morales ensured the researchers that the particular students in the study had only marginal differences in English proficiency. To address this as well, the assessment was designed in alignment with learning objectives that overlapped in each course. As noted earlier, the difference in criteria for higher-order thinking skills and author voice present additional limitations of the study. In particular, because performance on the criterion for higher-order thinking skills differed from task one to task two, more attention needs to be paid to how to

compare this aspect of the assessment in future studies. Finally, as this was our first use of the assessment approach, the purpose of this study was to seek evidence to see if this approach could provide dynamic evidence of student learning in this context. To accomplish this goal, we conducted a thorough analysis which also meant that the scores were not available to be used for grading in the course. Yet it is important to note that based on the findings from this study we are currently co-creating assessments with Mr. Morales which are being used for grading purposes in his courses.

### **Discussion**

In marrying heteroglossic perspectives about language with a flexible language assessment framework rooted in task-based language teaching and assessment, our goal was to not only create a valuable assessment tool, but to also transform the ways in which language teaching and learning positions multilingualism in assessment. Linguistically diverse populations often take language assessment that are used to make decisions about educational and vocational opportunities. In Oaxaca, Mexico, for example, our participants needed to pass classroom-based assessments to continue through the degree program toward graduation. The findings from this project suggest that this approach is promising because they are consistent with the general understanding that students tend to do better when their full linguistic repertoire is utilized. Specifically, the findings presented here serve to counter the view that students' multilingualism is irrelevant to their capacity to perform in the target language and point to potential implications for assessment of learners' higher-order thinking skills, as illustrated in our analysis of criterion four.

Our project integrated multilingual assessment approaches that aligned with the expressed needs of Mr. Morales while also taking into consideration some of the contextual

constraints of shared languages in the classroom. In doing so, the PAR methodological approach we employed pushes against attempting to attain an idealized fidelity of the assessment alignment with heteroglossia. We used an approach which was collaborative in order to address contextual necessities. In our study, meeting these contextual necessities meant that the assessments asked learners to write in English (i.e., students were not permitted to translanguage or use Spanish and English in their written responses). Thus we operated under assumptions that heteroglossia does not deny the existence of standardized varieties, nor the role of these languages within classrooms. Given the official curriculum requirements, such as the use of the TOEFL for graduation, we would argue that this approach is practicable even within a system that prioritizes standardized English. The findings in this study are seen as beginning to support more innovative and experimental approaches with multilingualism in language assessment. Ongoing collaborations with Mr. Morales are increasingly perforating the entrenched ideologies of English-only in assessments for pre-service English teachers in Oaxaca.

Our findings pointed to not only an improved performance when using readings in English and Spanish as opposed to English-only, but also may hold implications for performance on higher-order processing skills. In understanding that individuals may be able to demonstrate depth in the range of cognitive skills when assessments are inclusive of one's full linguistic repertoire, we view our study as a providing promising direction to conceptualize and integrate multilingualism as an asset in developing assessment approaches for linguistically diverse communities that value the language ecology of the context.

### **Conclusions**

We argue that these findings support further work in the development of classroom-based language assessments that value and sustain learners' language resources. This study offers

promising evidence that assessments can be developed with this approach and yield comparable and higher scores than those that favor monolingual approaches. Educational Linguistics framing of language practices as heteroglossic (Bakhtin 1981) and learning from a funds of knowledge perspective (González, Moll, and Amanti 2006)—brought together through continua of biliteracy lens in our assessment criteria (Hornberger and Link 2012)—will continue to serve as informative theories as we develop this task-based assessment approach in the future. Because what we presented here strengthens the case for language assessments that value and sustain learners’ language resources, in moving forward with this ongoing project we plan to delve deeper into having the linguistic diversity of the teachers and learners’ communities integrated into the assessment approach by guiding teachers in developing new assessments for their classrooms.

## References

Author 1 2014

Author 1 2015

Author 3 2013

Bachman, L., and A. Palmer. 2010. *Language Assessment in Practice*. Oxford, England: Oxford University Press.

Baker, C. 2003. “Biliteracy and Transliteracy in Wales: Language Planning and the Welsh National Curriculum.” In N. Hornberger (Ed.), *Continua of Biliteracy: An Ecological Framework for Educational Policy, Research, and Practice in Multilingual Settings* (pp. 71–90). Clevedon, UK: Multilingual Matters.

Bakhtin, M.M. 1981. *The Dialogical Imagination*, edited and translated C. Emerson and M. Holquist, Austin: University of Texas Press.

- Blackledge, A., & A. Creese 2010. *Multilingualism: A critical perspective*. Bloomsbury Publishing.
- Borda, O. F. 2001. Participatory (action) research in social theory: Origins and challenges. In *Handbook of action research: Participative inquiry and practice*, 27-37.
- British Council. 2015. "English in Mexico: An Examination of Policy, Perceptions and Influencing Factors." *British Council Report*.
- Canagarajah, A.S. 2013. *Literacy as Translingual Practice: Between Communities and Classrooms*. New York: Routledge.
- Cruz Martínez, A. 2016. "En México, Cada Vez Más Niños Padecen Diabetes Tipo 2," La Jornada. Retrieved May 16, 2017 from <http://www.jornada.unam.mx/2016/11/14/sociedad/041n1soc>.
- de la Luz Reyes, M. 2012. "Spontaneous Biliteracy: Examining Latino Students' Untapped Potential." *Theory into Practice* 51 (4): 248-255.
- Dewey, J. 1938/1997. *Education and Experience*. New York. Touchstone.
- Dodson, C. J. 1985. "Second Language Acquisition and Bilingual Development: A Theoretical Framework." *Journal of Multilingual & Multicultural Development* 6, (5): 325-346.
- Ellis, R. 2003. *Task-based Language Learning and Teaching*. Oxford, England: Oxford University Press.
- Freire, P. 2000. *Pedagogy of the Oppressed*. London: Bloomsbury Publishing.
- García, O. 2009. *Bilingual Education in the 21<sup>st</sup> Century: A Global Perspective*. Malden, MA: Wiley-Blackwell.
- García, O. & T. Kleyn. 2016. "Translanguaging Theory in Education." In *Translanguaging with Multilingual Students: Learning from Classroom Moments*, O. García and T.

- Kleyn eds., 9-33. New York: Routledge.
- González, N., L.C. Moll, and C. Amanti., eds. 2006. *Funds of Knowledge: Theorizing Practices in Households, Communities, and Classrooms*. New York: Routledge.
- Gwet, K. L. 2016. "Testing the Difference of Correlated Agreement Coefficients for Statistical Significance." *Educational and Psychological Measurement* 76 (4): 609-637.
- Higgins, C., R. Nettell, G. Furukawa, and K. Sakoda. 2012. "Beyond Contrastive Analysis and Codeswitching: Student Documentary Filmmaking as a Challenge to Linguicism in Hawai'i." *Linguistics and Education* 23 (1): 49-61.
- Hornberger, N.H. 1989. Continua of Biliteracy. *Review of Educational Research* 59: 271-96.
- Hornberger, N.H. 2001. "Educational Linguistics as a Field: A View from Penn's Program on the Occasion of its 25th Anniversary." *Working Papers in Educational Linguistics* 17 (1-2): 1-26
- Hornberger, N.H., ed., 2003. *Continua of Biliteracy: An Ecological Framework for Educational Policy, Research and Practice in Multilingual Settings*. Clevedon, UK: Multilingual Matters.
- Hornberger, N.H., and H. Link. 2012. "Translanguaging and Transnational Literacies in Multilingual Classrooms: A Biliteracy Lens." *International Journal of Bilingual Education and Bilingualism* 15 (3): 261-278.
- Hornberger, N.H., and E. Skilton-Sylvester. 2000. "Revisiting the Continua of Biliteracy: International and Critical Perspectives. *Language and Education: An International Journal* 14: 96-122.
- Horner, B., M. Lu, J.J. Royster, and J. Trimbur. 2011. "Opinion: Language Difference in Writing--Toward a Translingual Approach." *College English* 73 (3): 303-321.

- Hult, F.M. 2007. "The History and Development of Educational Linguistics." In B. Spolsky and F.M. Hult (eds.), *The Handbook of Educational Linguistics* (10-24). Malden, MA: Blackwell.
- Ibrahim, A. 2008. "Operating Under Erasure: Race/language/identity." *Comparative and International Education/Éducation Comparée et Internationale*, 37 (2): 56-76.
- Jain, R. 2014. "Global Englishes, Translinguistic Identities, and Translingual Practices in a Community College ESL Classroom: A Practitioner Researcher Reports." *TESOL Journal* 5 (3): 490-522.
- Jørgensen, J.N. 2008. "Polylingual Languageing Around and Among Children and Adolescents." *International Journal of Multilingualism* 5 (3): 161-176.
- Jørgensen, J., L. Karrebaek, M. Madsen, and J. Møller. 2011. "Polylanguageing in Superdiversity." *Diversities* 13 (2): 23-38.
- Krippendorff, K. 2011. "Agreement and Information in the Reliability of coding." *Communication Methods and Measures* 5 (2): 93-112.
- Lado, R., 1979. "Thinking and 'Languageing': A Psycholinguistic Model of Performance and Learning." *Sophia Linguistica* 12: 3-24.
- Lewis, G., B. Jones, and C. Baker. 2012. "Translanguageing: Origins and Development from School to Street and Beyond." *Educational Research and Evaluation* 18 (7): 641-654.
- Lin, A. M., and Y. Wu. 2015. "'May I speak Cantonese?'-Co-constructing a Scientific Proof in an EFL Junior Secondary Science Classroom." *International Journal of Bilingual Education and Bilingualism* 18 (3): 289-305.
- Mohan, B.A. 1986. *Language and Content*. (Vol. 5288). Addison Wesley Publishing Company.

- Mohan, B.A. 1990. "LEP Students and the Integration of Language and Content: Knowledge Structures and Tasks." *Proceedings of the Research Symposium on Limited English Proficient Students' Issues*. Washington, DC.
- Makoni, S., and A. Pennycook. 2005. "Disinventing and (Re)Constituting Languages." *Critical Inquiry in Language Studies: An International Journal* 2 (3): 137-156.
- McIntyre, A. 2008. *Participatory Action Research*. Thousand Oaks, CA: Sage Publishing.
- Moll, L.C., C. Amanti, D. Neff, and N. González. 1992. "Funds of Knowledge for Teaching: Using a Qualitative Approach to Connect Homes and Classrooms." *Theory into Practice* 31 (2): 132-141.
- Norris, J.M. 2016. "Current Uses for Task-Based Language Assessment." *Annual Review of Applied Linguistics* 36: 230-244.
- Otheguy, R., O. García, and W. Reid. 2015. "Clarifying Translanguaging and Deconstructing Named Languages: A Perspective from Linguistics." *Applied Linguistics Review* 6 (3): 281-307.
- Pacheco, M.B., and B.E. Smith. 2015. "Across Languages, Modes, and Identities: Bilingual Adolescents' Multimodal Codemeshing in the Literacy Classroom." *Bilingual Research Journal* 38 (3): 292-312.
- Palmer, D.K., R.A. Martínez, S.G. Mateus, and K. Henderson. 2014. "Reframing the Debate on Language Separation: Toward a Vision for Translanguaging Pedagogies in the Dual Language Classroom." *The Modern Language Journal* 98 (3): 757-772.
- Paris, D. 2012. "Culturally Sustaining Pedagogy: A Needed Change in Stance, Terminology, and Practice." *Educational Researcher* 41 (3): 93-97.



- Pica, T. 2005. "Classroom Learning, Teaching, and Research: A Task-based Perspective." *The Modern Language Journal* 89 (3): 339-352.
- R Core Team, 2016. "R: A language and environment for statistical computing." *R Foundation for Statistical Computing, Vienna, Austria*. <https://www.R-project.org/>.
- Ramírez Victoria, G.A. and E.N. León Jiménez. 2012. "Plad de Desarrollo de Unidad Académica 2012-2015." *Facultad de Idiomas Report*.
- Rea-Dickens, P., G. Yu, O. Afitska, F. Olivero, N. Ingram, S. Erduran, Z. Khamis, A. Mohamed, M. Mbaraok, H. Mwevura, and S. Said. 2011. "Investigating the Language Factor in School Examinations: Exploratory Studies." SPINE Working Papers (2). Bristol: University of Bristol.
- Shohamy, E. 2011. "Assessing Multilingual Competencies: Adopting Construct Valid Assessment Policies." *The Modern Language Journal* 95 (3): 418-429.
- Shohamy, E. 2006. *Language policy: Hidden agendas and new approaches*. New York: Routledge.
- Spolsky, B. 1978. *Educational Linguistics: An Introduction*. Rowley, MA: Newbury House.
- Stansfield, C.W. 2011. Oral Translation as a Test Accommodation for ELLs. *Language Testing* 28 (3): 401-416.
- Swain, M. 2006. "Languaging, Agency and Collaboration in Advanced Second Language Proficiency." *Advanced language learning: The contribution of Halliday and Vygotsky*: 95-108.
- Vaish, V., and A. Subhan. 2015. "Translanguaging in a Reading Class." *International Journal of Multilingualism* 12 (3): 338-357.
- Van Gorp, K., and B. Deygers. 2014. "Task-Based Language Assessment." *The Companion to*

*Language Assessment*. DOI: 10.1002/9781118411360.wbcla058

Vazquez, A.M., N.P.T. Guzmán, and R. Roux. 2013. “Can ELT in Higher Education Be Successful? The Current Status of ELT in Mexico.” *TESL-EJ*, 17 (1).

Whyte, W.F.E. 1991. *Participatory Action Research*. Thousand Oaks, CA: Sage Publications, Inc.

Zapf, A., S. Castell, L. Morawietz, and A. Karch. 2016. “Measuring Inter-rater Reliability for Nominal Data—Which Coefficients and Confidence Intervals are Appropriate?.” *BMC Medical Research Methodology* 16 (1):93: 1-10.

## Appendix A

### Task one: Advice (Spanish/English)

1. Learning objective from syllabus: Unit 1 on health and writing objective “Write a letter or an email to a friend in order to give advice on how she/he can change her/his habits in order to be healthier.”
2. Language function: Giving advice (suggesting action)
3. Skill: annotation and synthesizing from multiple sources
4. Modes: reading and writing
5. Audience: informal register
6. materials: three readings
  - a. Narrative of experience with diabetes (English)
  - b. Statistics of diabetes in Mexico (English)
  - c. Information about preventative health care for diabetes (Spanish)

#### Assessment task objectives:

Participants will **annotate** readings by underlying key words or making comments. Participants will **synthesize** sources in Spanish and English about diabetes to give advice by writing an email to a friend.

### Task two: Recommendation (English)

1. Learning objective from syllabus: Unit 2 on societal problems and the economy objective “Write a letter to a community leader complaining about a problem in your neighborhood.” And “Write a description of a very important problem in the country and how to solve it.”
2. Language function: Proposal (suggesting actions)
3. Skill: annotation and synthesizing from multiple sources
4. Modes: reading and writing
5. Audience: formal register
6. materials: three readings
  - a. Statistical information about garbage in Mexico (English)
  - b. Recommendation option 1: Billboard punishment (English)
  - c. Recommendation option 2: Adopt-a-street program (English)

#### Assessment task objectives:

Participants will **annotate** readings by underlying key words or making comments. Participants will **synthesize** sources in English about litter to make recommendations about a litter program by writing a letter to university administration.

*Table 1*  
*Rubric Criteria for Task Completion*

Rubric criteria	Task 1	Task 2
Interaction with texts	Reading annotations	Reading annotations
Context	Diabetes in Mexico	Garbage in Mexico
Comprehension	Two pieces of advice	Recommendation for campus program
Higher-order processing	Analyze: Cause and effect	Evaluate
Audience	Email informal tone	Letter formal tone
Author voice	Personal perspective	Personal perspective

*Table 2*

*Assignment of Participant Tasks by Rater*

Task one		Task two	
Rater A	2-22	Rater E	2-22
Rater B	23-44	Rater F	23-44
Rater C	2-11; 23-34	Rater G	2-11; 23-34
Rater D	12-22; 35-44	Rater H	12-22; 35-44
Training participants: 1, 19, 27		Training participants: 1, 27, 32	
Excluded participant: 32		Excluded participant: 19	

*Table 3*

*Results of T-test and Descriptive Statistics Multilingual TBLA*

Outcome	Task 1		Task 2		n	95% CI for Mean Difference	t	df
	M	SD	M	SD				
	4.654	0.912	4.077	0.899	39	0.225, 0.929	3.32*	38

\*  $p < .05$ .

Table 4

*Results Comparing Paired T-tests and Descriptive Statistics*

Criteria	Task 1		Task 2		n	$\hat{d}$	95% CI for Mean Difference	t	df
	M	SD	M	SD					
All (1-6)	4.65	0.91	4.08	0.90	39	0.63 <sup>a</sup>	0.23, 0.93	3.32*	38
Criterion 4 excluded	4.05	0.88	4.01	0.84	38	0.05 <sup>a</sup>	-0.29, 0.37	0.24	37

*Note.* Cases where the difference between two rater scores was 2 or more points were removed from the analysis. When all 6 criteria were used, one of 40 cases was removed (n = 39). When criterion 4 was excluded, two of 40 cases were removed (n = 38).

<sup>a</sup> Cohen's d was estimated using the standard deviation of Task 2 scores.

\* p < .01.